AISec: Leveraging Artificial Intelligence for Personalized Security and Privacy

A white paper in support of the NITRD RFI on the Five-Year Strategic Plan for the Federal Networking and Information Technology Research & Development Program

There is a long tradition of using artificial intelligence (AI) to tackle security problems. A prevalent research method is to collect data capturing a particular malicious activity (e.g. network intrusions, spam) and using AI techniques such as machine learning, train a detector for future malicious activity of the same type. While this approach clearly yields powerful security technologies it is largely an after-the-fact approach to security in that data of security breaches is needed first, before the protection mechanism can be developed. The growth of the Web and efficient AI-based techniques for mining large corpora mean that we can now *anticipate* the adversary to a degree not possible previously. While this is a significant advance, we argue that even more transformative technologies are possible through continued collaboration between AI and security. In particular, we call for research in a new subdiscipline called "AISec" that leverages and extends AI advances in predictive modeling to achieve *personalized* security technology. We envision security technology that is personalized to the user in terms of their security vulnerabilities and privacy preferences, thus achieving high usability while providing strong security and privacy.

We highlight three problem areas which have recently gained a proactive security advantage through leveraging AI before discussing the AISec agenda in more detail.

Password reset. Online service providers commonly require users to not only select "challenge questions" that they may be asked in the event that they forget their password and are unable to login. Common challenge questions include, "What's your mother's maiden name?" and "What's your favorite pet's name?". Recently, Jakobsson et al demonstrated that publicly available records can be mined to determine the answers to many such questions (see, for example, [GJ05]) . Subsequently, they developed a more secure approach based on preferences [JSWY08].

CAPTCHAs. The tests humans are asked to perform when registering for a Web site (typically, typing in a distorted word) are called CAPTCHAs. The state of the art in CAPTCHAs is constantly evolving as advances in computer algorithms frequently ruin the effectiveness of a particular CAPTCHA at distinguishing between humans and computers. Despite this "arms race", CAPTCHAs are routinely presented and put into use with little, if any, testing, often with bad results. For example, in [G08], Golle demonstrated that standard machine learning techniques can be used to break a CAPTCHA introduced at one the security community's most competitive conferences, thus underscoring, that AI techniques should be routinely applied to proactively identify weaknesses.

Data Privacy. Documents are commonly redacted prior to release to protect sensitive content when responding to Freedom of Information Act (FOIA) requests or legal subpoena. While the act of redaction can be time-consuming and tedious, the process of determining what to redact is even more challenging and error-prone (examples of redaction failures are discussed in [SGZ07]). In [CGS08] a fast data mining approach was demonstrated that allows the keywords closely associated with a sensitive topic to be quickly identified. The approach leverages the Web to model the adversary's knowledge and provide proactive privacy protection against inferences.

An Opportunity: Personalized Security.

While the above technologies are novel and effective, we believe even more transformative technologies are possible through a closer collaboration between the AI and security and privacy communities. The data mining community has demonstrated the power of predictive models for advertising. Even seemingly generic data like browser history or search

terms has proven to be strongly indicative of demographic attributes (see, for example [adLabs]) and even identity [NYT]. We propose that the AI and security communities work together to design analogous models to transform the way users experience security and privacy today. With a sharing of data between industrial partners, the key features to predictive models of a user's security habits and privacy preferences can be identified. Such models will enable the user's security experience to be tailored to them, that is, their security vulnerabilities and privacy preferences, thus increasing usability while providing better security and privacy.

Consider for example, a model based on the types of software applications installed, browser habits and history, keyword searches and network connection patterns. The model might suggest that someone who frequently clicks on links in emails, connects to many unknown wireless networks and uses short passwords is a risk-taker and so needs a more stringent security policy. In particular, risk-takers might experience stricter browser requirements around certificate acceptance, and less flexible security posture requirements for connection to their employer's internal network. In contrast, users with good security practices, might enjoy more leeway with installing security patches and fewer hurdles to corporate intranet access.

Similarly, a user who engages in limited online social networking and contributes anonymously when they do, might be predicted to have strong privacy concerns around demographic information. For such a user, relevant parts of a Web site's privacy policy could be highlighted for them before they register, and they could be warned about sites that are likely to violate their privacy preferences. This automatic prediction of privacy preferences would be especially powerful in light of the well-documented difficulty users have in articulating their true privacy concerns [LHDL04].

Building on work done in the AI community on social influence (see, for example, [CCHS08]) the model might also predict the user's vulnerability to social engineering attacks. A vulnerable user could be supported with a more stringent warning system or automated protections. For example, an automated protection mechanism might bounce suspicious emails that appear to come from friends and send re-send requests to friends using email addresses from the recipient's contact list.

In conclusion, it is well-documented that one-size-fits-all security mechanisms frequently frustrate users and as result are often simply turned off, resulting in no security at all (see, for example, [WSC06]). We believe a research agenda in *personalized* security would translate into less corporate and government data leaks, as it targets security policy for the user and incentivizes good behavior. The interdisciplinary nature of this research agenda and the requirement for collaboration between multiple industrial partners and academia makes support from an influential government body like the NITRD crucial.

[adLabs] Microsoft's adCenter Labs. http://adlab.msn.com/
[CGS] *Detecting Privacy Leaks Using Corpus-Based Association Rules*. R. Chow, P. Golle and J. Staddon. KDD 2008.
[CCHS08] *Feedback Effects between Similarity and Social Influence in Online Communities*. David Crandall, Dan Cosley, Daniel Huttenlocher, Jon Kleinberg, Siddharth Suri. KDD 2008.
[G08] *Machine Learning Attacks against the Asirra CAPTCHA*. P. Golle. ACM CCS 2008.
[GS05] *Messin' with Texas, Deriving Mother's Maiden Names Using Public Records*. V. Griffith and M. Jakobsson. ACNS '05.
[LHDL04] *Personal Privacy through Understanding and Action: Five Pitfalls for Designers*. Lederer, S., J.I. Hong, A. Dey, and J.A. Landay. Personal and Ubiquitous Computing 2004. 8(6): p. 440 - 454.
[JSWY08] *Love and Authentication*. M. Jakobsson, E. Stolterman, S. Wetzel, L. Yang. (Notes) ACM Computer/Human Interaction Conference (CHI), 2008.
[NYT] *A Face Is Exposed for AOL Searcher No. 4417749*. M. Barbaro and T. Zeller. New York Times, August 9, 2006.
[SGZ] *Web-Based Inference Detection*. J. Staddon, P. Golle and B. Zimny. USENIX Security 2007.
[WSC06] *User experiences with sharing and access control*. T. Whalen, D. K. Smetters. E. Churchill. CHI Extended Abstracts 2006: 1517-1522